

THE ANATOMY OF INNOVATION

*

How do major technological changes occur in human societies? In what circumstances are existing products and practices set aside in favour or radically new ones? Just what is it that makes a particular new fashion at a specific time displace an old one?

Innovation, the development or introduction of what is new, is evidently a process whose understanding is fundamental to the study of society and especially of change. It is one which archaeology has yet to deal with successfully. The purpose of this paper is twofold. First to argue that the recent popularity of 'diffusionist' explanations in archaeology, and the undoubted success of modern geographers in studying innovation diffusion as a spatial process, have obscured an important central point: that in many cases a spatial framework is not the most appropriate for the study of innovation. Secondly, to offer an alternative model where the crucial mechanism is human choice – the conscious decision by the individual to adopt one mode of undertaking a particular activity rather than another. In many cases human choice is bimodal – a yes/no decision between two alternatives. Catastrophe theory may be used to investigate the properties of discontinuity, hysteresis and divergence which accompany bimodal change. The rapid and widespread appearance of an innovation in the archaeological record need not indicate the operation of some external agency upon the culture system.

These problems of invention and adoption are inextricably linked in much archaeological literature on the subject, yet I believe that they can profitably be separated, since in some cases very different processes are at work. A good example is offered by copper metallurgy, first seen in the Old World at a very early date in the Near East. Yet a productive copper technology with a large output de-

veloped only millennia later in that area, by which time indications of the necessary technical skill to smelt and cast copper are evident in south-east Europe and perhaps in Iberia. That this circumstance should not automatically be seen as the working of 'diffusion' has been argued in a number of places (Renfrew 1969, Tringham 1971, 198) and refuted in others (Wertime 1964, 1257). Some rather more careful analysis is desirable.

Change in Society

Innovation is to be distinguished from invention. Invention is the discovery or achievement by an individual of a new process or form, whether deliberately or by chance. Innovation as it will be understood here implies the widespread *adoption* of a new process or form, and clearly it must be preceded by the relevant inventions whether by a short or by a long period. 'In the usual terminology the inventions, small and large, needed for progress become "innovations" when entrepreneurs adopt them in industry. The innovational process requires of course not only the production but the distribution of additional knowledge, and it requires entrepreneurial decisions' (Fellner 1971, 2). Whether we are speaking of industrial or domestic activity it is essential to distinguish between *invention* and *adoption*.

The initial knowledge required to carry out the process in question may have been gained locally by invention or may have been introduced into the area in question as a result of contacts with another area where the invention was earlier made. This process is frequently termed 'diffusion' by archaeologists although the mechanism of transmission of the knowledge is often not defined. Sometimes the knowledge transmitted may be only partial: contact may impart some general knowledge of the outlines of a process or form without its precise details, which are re-invented anew under the stimulus of the contact. This is what Kroeber (1940) calls 'stimulus diffusion', and he gives as examples the development of fine ceramics in western Europe under the influence of Chinese porcelain, and the invention of a written syllabary for the Cherokee language after contact with the European writing system. But Kroeber does not make the distinction emphasised here: only the first of his examples ranks as an innovation. The second is merely an invention, since the Cherokee writing devised by Sequoya did not become widely used.

In the same way the circumstances favourable to the adoption of a new process may develop entirely locally, or they may come about

through continuing contacts with another society, whether or not the initial invention was learnt from that source. Childe's model for the development of bronze age society in central Europe under the influence of a developing trade with the Mycenaean world (Childe 1958) is an example of the latter kind. Analogous mechanisms are implied in Fried's distinction between 'pristine' and 'secondary' civilisations (Fried 1967), the essential innovations of the latter being assumed to derive from the former.

Archaeologists in recent years have rightly stressed the importance of the adoption process, and the circumstances in the social context which favour it. As Flannery (1968, 80) puts it,

The use of a cybernetics model to explain prehistoric cultural change, while terminologically cumbersome, has certain advantages. For one thing it does not attribute cultural evolution to 'discrepancies', 'inventions', 'experiments' or 'genius', but instead enables us to treat prehistoric cultures as systems . . . it allows us to view change not as something arising *de novo*, but in terms of quite minor deviations in one overall part of a previously existing system, that once set in motion can expand greatly because of positive feedback.

The systems approach does indeed offer a framework whereby the adoption of an invention can be examined. Positive feedbacks linked to give a 'multiplier effect' (Renfrew 1972, 27-44; chapter 9 above) may be used to explain gradual changes. But they do not so readily explain the sudden innovation, the rapid adoption, which is often taken as an indication of external intervention, and that is where catastrophe theory can help.

Diffusion as a Spatial Process

Radical changes in the productive technology and the organisation of human society are often described and investigated in terms of 'diffusion' – that is to say within an explanatory framework where spatial co-ordinates have a special significance. It is important to realise, however, that the underlying assumptions, rarely stated by archaeologists, do not always hold. Moreover there are arguments for suggesting that these assumptions are valid less often for prehistoric times than for the twentieth century A.D.

It should be noted also that the term 'diffusion' is used in some contexts in the sense of 'migration'. Chemists speak of diffusion of molecules across a membrane, for instance. And some geographical analyses have found that similar models are effective whether one is speaking of the introduction of a new population into an empty area

or the spread of an invention in an already existing human population. By diffusion we mean a process in which very few human individuals are displaced by more than a few kilometres.

One of the most successful diffusion models for the mechanism of culture change is the 'infectious' or 'genetic' explanation. Here the progressive adoption of a new process in a human community is analysed spatially. The underlying postulate or assumption is that the new process can only spread by contact with a carrier – someone himself already 'infected' with the innovation, someone already carrying the crucial gene. The law of transference by personal contact usually brings as a consequence a spatial continuity in the dispersal of the observed innovation (unless an infected individual travels rapidly to a new location and starts fermenting infection there). The basic mechanism of the culture process is infection through contact (although it need not be assumed that 100 per cent of those exposed to infection in fact catch the innovation). The role of the receiving individual (or culture) is entirely passive – he may have a high or low resistance to infection, be an 'early adoptor' or a 'late acceptor' – but his personal contribution is about as significant as that of a guinea pig exposed to a virus.

In geography the simulations of Hägerstrand and his colleagues have indeed illuminated significantly the processes of spatial diffusion of technological improvements in recent Europe as well as the movements of settlement. Hägerstrand's *Innovation Diffusion as a Spatial Process* (1967) is one of the most important works of the past decade, for the archaeologist as well as the geographer, and introduces several refinements in terms of differing receptiveness and effectiveness of communication which make it strikingly apposite for the cases studied.

It is not my purpose to doubt the value of such models here, but simply to present an alternative perspective which may, in certain cases, be preferable. It should be stressed, moreover, that these models are formalisations of a general outlook which has been widespread and too little questioned within archaeology for more than a century.

Most diffusion models carry with them, or rather with their application, two fundamental assumptions which, if applied very generally, would between them exclude most of the fundamental workings of culture process:

- 1) Change is exogenous to the location in question.
- 2) Widespread adoption of a new process follows rapidly in a regular manner after exposure to 'infection' (i.e. knowledge of the

invention), allowing for different degrees of receptiveness among those infected, and for various delays related to efficiency of communication.

The first assumption if applied over very wide areas carried with it the logic of Childe's splendidly 'useful heuristic hypothesis that each innovation has been made but once.' (Childe 1956, 154). The view that the fundamental inventions only occur once in human history is a basic tenet of doctrine among some students of the history of technology, for instance among many metallurgists:

One must doubt that the tangled web of discovery, comprehending the art of reducing oxide and the sulfide ores, the recognition of silver, lead, iron, tin and possibly arsenic and antimony as distinctive new metallic substances, and the technique of alloying tin with copper to make bronze, could have been spun twice in human history. (Wertine 1964, 1257)

The second assumption fails to take into account that the basic technology required for the adoption of many new processes is often available, whether through local development or as a result of contact with neighbouring societies, decades or centuries before it is in fact utilised on a large scale.

These comments are not intended as adverse criticism of Hägerstrand's work, and I do not suppose that he would disagree with them. In his discussion of the characteristics of innovation diffusion (1967, 133-4) he notes what can be regarded as spatial regularities of the cases under study:

Stage 1. Local concentrations of initial acceptancy;

Stage 2. Radial dissemination outward from the initial agglomerations, while those original centres simultaneously continue to condense;

Stage 3. The growth ceases. He explicitly states that 'the model population is imagined to be in the process of adopting a cultural innovation of exogenous origin.' (*ibid.*, 138).

In the first of his proposed models, which does not prove effective, it is assumed (a) from the beginning the entire population is informed about the new process, and (b) acceptances occur independently of one another in random order of precedence. In his second model (a) in the beginning only one person in the population is informed about the new process and (b) acceptance occurs immediately upon the receipt of information. In his third model: (a) only one person in the population has accepted the fictional new process from the beginning, and (b) acceptance does not occur until resistance (according to the given scale) has been overcome through

repeated receipt of private information from persons who previously adopted the invention. In the second and third models, therefore, which are those which Hägerstrand develops, the crucial process is that of exposure to information about the new process.

The two basic assumptions indicated above do in fact hold for some important cases, and their success with these has obscured their unsuitability in others. In the first place they are highly relevant when one is dealing with an introduced species new to the area. You cannot have cereal farming without cereals, or silk without silkworms, any more than bubonic plague without a bacillus. Until genetic engineering develops further, assumption (1) is in fact true for all areas but the homeland of the exotic species in question. Assumption (2) may well hold also where the new species or variety fits easily into the existing agricultural system. (The model of Ammerman and Cavalli-Sforza (1973), although based on ideas of epidemiology, works because of the increase in population density which they see as associated with the inception of agriculture, and it does not rest on assumption (2).)

The assumptions are appropriate also for certain technological advances in the modern world, such as those which Hägerstrand discusses, where the positive benefits of the innovation are unquestionably very substantial. The case is not considered (except briefly, in a regional context) that the invention may be of benefit to some individuals but not to others according to circumstance, nor that the benefits themselves may vary through time.

Prior to the Industrial Revolution, invention did not take place primarily in Research Institutes or in Research and Development Laboratories within a framework of pure science, but rather was the result of practical men discovering improvements in the course of their craft activities. Discovery and invention were the work of craft specialists rather than pure scientists. The opportunity for invention was thus present wherever specialist crafts were practised: there were more potential primary centres. It may be reasonable for Hägerstrand (1967) to assume that tuberculin testing was exogenous to rural Sweden, but the principle of exogenous origin cannot be readily assumed by prehistorians.

I would argue moreover that many early inventions or discoveries, for example the exploitations of domesticable plants or the widespread manufacture of copper tools, only became profitable and hence viable as innovations with gradual change in a number of circumstances. Assumption (2), appropriate enough as used by Hägerstrand, cannot reasonably be accepted prior to the Industrial

Discontinuity and Long-Term Change

Revolution, except for a limited number of inventions. Indeed it applies directly only to those inventions which allow an existing process to be carried out more cheaply and efficiently without any compensating disadvantages, and without substantial capital cost.

In order to construct a clear, alternative model, let us reverse these two diffusionist assumptions, replacing them by two others:

(i) Any functional invention or innovation in human technology or culture will ultimately occur again, and recurrently, irrespective of time, space or ethnic group, given the appropriate conditions.

(ii) Widespread adoption of a new process does not follow automatically upon the inception (whether locally or by outside contact) of the new process or form but depends in a complicated way upon individual choice governed by social and other factors. Availability of the technical means is a necessary but not a sufficient condition for the adoption of new inventions or discoveries.

It will be argued below that the early origin of copper metallurgy in Europe offers an excellent example for the development of these alternative assumptions. The technical processes of smelting and casting of copper are crucially dependent on the local availability of raw material and fuel, and the pyrotechnical skills to handle these materials. Both these conditions were fulfilled in the Balkans at the inception of the Copper Age and in the Iberian peninsula at the onset of the Chalcolithic, copper technology following in both cases upon ceramic technology. We see, therefore, the early handling of copper and other metals, in several areas with a settled, farming population and developed ceramic technology. Sometimes, as in the fourth millennium in the Aegean area, the technical skills were available for up to a millennium before the invented process was widely adopted, with a developed metal age as consequence.

It is necessary therefore to contrast the rather naive (although often strikingly effective) simplicity of the 'infection' model (figure 1) with the more careful consideration of the 'innovation choice' model (figure 2).

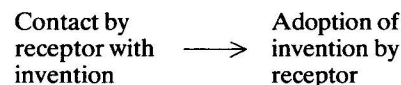


Figure 1. Innovation by infection/diffusion.

Under different conditions each will be appropriate, and it is not my purpose to question the usefulness of the former, simply its universality. It is in fact a special case of the latter. The focus of our

The Anatomy of Innovation

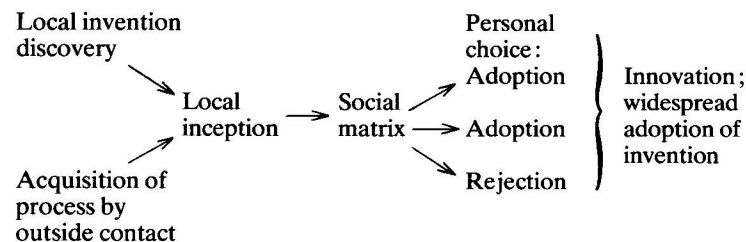


Figure 2. Innovation through contextual choice.

interest will more often be upon the mechanisms governing *choice* in the society, and not upon those of spatial contact.

Catastrophe Theory and Bimodal Choice

Change in human society often occurs suddenly. Settlement patterns alter rapidly, centralised organisations undergo rapid collapse, and new processes are widely adopted within a short space of time. In these circumstances, not surprisingly, the archaeologist and the historian have frequently sought some equally striking cause – something new and influential to explain the rapid change, sometimes seen as an apparent discontinuity in the archaeological record. The obvious first choice as an explanation for the observed discontinuity has often been a discontinuity in the circumstances resulting from an invasion or migration. In the absence of evidence for these, the obvious second choice has been decisive influence from outside the society in question to account for the radical change.

One of the insights offered by catastrophe theory is that striking discontinuities in the behaviour of systems are to be expected, under certain circumstances, from the smooth and perfectly continuous change in the conditions which govern behaviour. Thus the collapse of early state societies, and in some circumstances their sudden formation, can be described without the intervention of exogenous forces. Sudden shift in settlement pattern is to be predicted, given certain gradual changes operating entirely within the culture system. As we shall see, innovation, the widespread acceptance of a new process in society can take place quite suddenly, through the local operation of relevant factors, changing quite smoothly. Whether the technical process underlying the innovation was ultimately of local or outside inception can be quite immaterial to the analysis.

Although the argument here was prompted by recent applications of catastrophe theory in various fields (Thom 1975, Zeeman 1976), the discussion does not require us to use any very advanced mathematics. Moreover the cusp catastrophe, the only one of Thom's 'magnificent seven' elementary catastrophes with which we have to deal, has already been described in chapter 13 and a number of accessible papers (Amson 1975, Isnard and Zeeman 1976, Renfrew 1978, Wagstaff 1976) and the description here can be brief.

We imagine a system whose behaviour may be described by a single variable, which we shall call the behaviour or state variable x . In the case we are to discuss, bimodal choice, x represents the behaviour of the individual who is doing the choosing. He is choosing between two options, one represented by positive values of x , the other by negative values. This is not, however, simply a yes/no decision but a strategy mix – for instance the percentage of two crops grown, or the householder wishing to decide how many gas and how many electrical appliances to use in his house, or the traveller deciding how many miles to travel by bus and how many by rail.

The behaviour is influenced by variation in two control variables, a and b . For instance a could represent relative speed of travel, and b comfort of the bus versus rail example – or in crossing the Atlantic by ship or aircraft; or a could represent the crop yield of a particular mix of crops, and b the labour input required to cultivate that mix.

We further imagine a potential function or utility function $f(a, b, x)$ whose value ϕ represents the perceived overall benefits or attractiveness of the strategy x for different values of a and b .

$$\phi = f(a, b, x)$$

We assume that the individual acts in such a way as to choose the value of x that is locally optimal (in terms of perceived benefits or attractiveness). Then for given values of a and b the possible equilibrium values of x are the maxima of ϕ , and hence will be solutions of

$$\frac{\partial \phi}{\partial x} = 0$$

As the control variables a and b vary, the values of x which are the maxima of the above equation determine a surface in the space (a, b, x) . In most cases such surfaces are single sheets, there being a single maximum (optimum) values of x for each point (a, b) in the control space defined by the control variables. But there may also

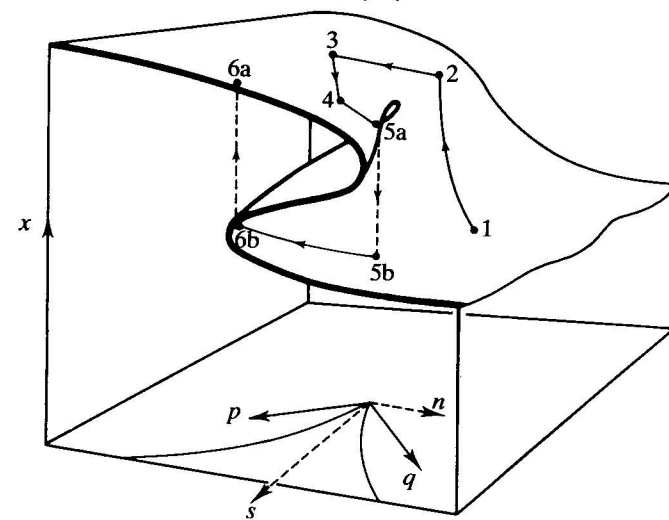


Figure 3. The cusp catastrophe, illustrating bimodal choice.

be values of (a, b) for which there exists more than one maximum of x , in which case the surface is folded. When there is more than one maximum value of x , the state variable, for given values of a and b , the control variables, a sudden change of state of the system may occur as x moves discontinuously from one maximum to the other. Such a sudden change is termed by Thom (1975) a 'catastrophe'. Thom's Theorem of Elementary Catastrophes, crudely paraphrased, assuming structural stability, states that if the number of control variables of such a system is not more than 5, any singularity (or specially twisted region of the map which underlies the discontinuities in observed behaviour) is equivalent to one of a finite number of types called elementary catastrophes, seven in number. When there are only two control variables the singularities can be of only two specific kinds, which Thom calls the *fold catastrophe* and the *cusp catastrophe*.

The cusp catastrophe is illustrated in figure 3. It represents the most complicated local behaviour possible when we have two control variables, and at the same time the simplest when we know that there is indeed a discontinuity which we wish to examine in which the state variable suddenly changes from one value to another.

The surface is given by the equation:

$$x^3 + ax + b = 0$$

On this surface the values of x , a and b are such that the utility function $f(a, b, x)$ is a maximum (or minimum, saddlepoint, etc.). The utility function itself is given by the equation:

$$f(a, b, x) = \frac{1}{4}x^4 + \frac{1}{2}ax^2 + bx$$

Yet the beauty of the approach is that this form need not in fact be known for a particular application. Thom's theorem states that, given his initial assumptions, there are only a few ways that discontinuous change can occur, corresponding to his seven elementary catastrophes. In this case with two control variables there are only two. The cusp catastrophe is appropriate for the consideration of bimodal choice.

Let us now consider a specific hypothetical case. x is the state variable representing the extent to which the strategy mix of the individual traveller under consideration favours rail rather than car travel (running for instance from -50% to $+50\%$ for all car to all rail respectively). The control variable p represents the speed for a given journey for rail as opposed to car, and q the cost for a given journey for rail as opposed to car.

The overall benefits ϕ of a given strategy mix x are seen in figure 4 for different values of p and q .

Figure 4a indicates the situation when p , the relative speed of rail travel, is high, and q , its relative cost, is low. The benefits are greatest (optimal) for a high value of x . A high percentage of the total miles travelled by the individual will be by rail (although not, of course, all, since not all his itinerary will be served by railways).

Figure 4b shows the case where the relative speed of rail to car p is low and the relative cost q is high. The benefits are greatest for low x , with most of the travel by car.

Figure 4c shows the case when both p and q are intermediate – costs of both rail and car are about the same, and so is speed. In such a case the optimum of x will take an intermediate value representing a balanced strategy between car and rail.

Figure 4d shows the case when p and q are both high. There are now two optimum values for x – one high (representing the rapidity of rail travel, despite its greater cost) and one low (representing the relative cheapness of car travel, despite the greater time required).

At this point we can appeal to Thom's theorem and state that if there are singularities, and given certain further assumptions of a mathematical nature, then the most complicated form the configuration can take is that of the cusp catastrophe (figure 3).

We are now in a position to grasp the complete anatomy of the

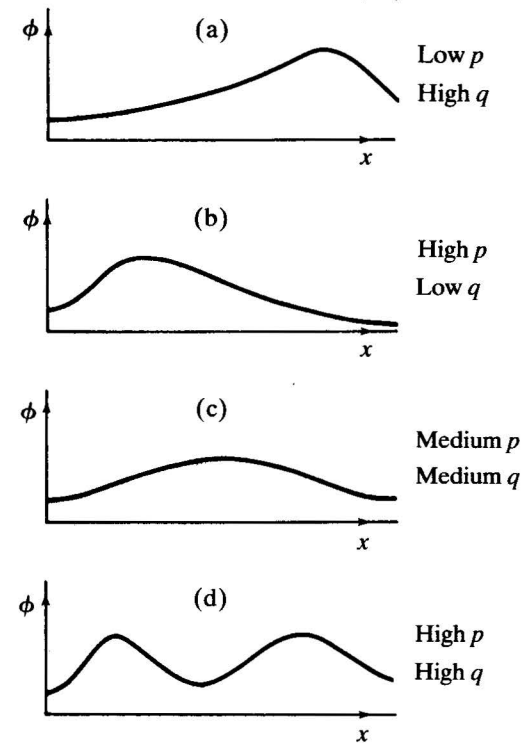


Figure 4. The value of the utility function ϕ against the measure for strategy choice x for different fixed values of the control factors (p , q).

situation. The surface seen in figure 3 represents the values of x for which the function ϕ is a maximum – it shows the value of x (for the given values of p and q) representing an optimising strategy in terms of both the variables. It therefore represents the values of x for which the system is in equilibrium. So as p and q change (for instance along the trajectory times t_1 to t_3 represented by points 1 to 3) so does the optimal value of x and hence the behaviour of the individual concerned.

If, for example, q is initially high and p low, car travel being cheaper and faster than rail, then the individual will wherever possible travel by car, and x is low (point 1). If petrol costs rise, reducing q , he may balance his strategy (point 2). But then rail travel is improved and becomes faster (point 3). At t_4 , however, the cost of rail travel is increased and the surface is two-valued in x for

the given values of p and q . Probably, being accustomed to rail travel, our individual will continue to prefer this, until an increase in car speeds (decrease of p) brings him to point 5a. By now the attractiveness of rail has completely waned – it is dearer and it is now no faster – so at this point he makes the conscious and firm decision to travel by car.

The switch at this point is a discontinuous one, and the point on the surface moves from 5a (high x) to 5b (low x). The individual has now switched his strategy, perceiving that the benefits are now in favour of car travel.

If rail speed relative to car should again rise (p rising) he will tend to stick to his existing strategy until point 6a is reached, when the proposition of travelling by car is no longer attractive: he then switches to rail (6b).

We shall return below to the question of transport choice (see also Ferguson 1976; Wilson 1976), but first it should be stressed that the above description of bimodal choice is of very great generality. In general, if x represents the choice or mix between two possible courses of action, and p and q are conflicting control variables, each acting to counter the effects of the other, the configuration will be precisely as described here. A further example is offered by Zeeman's dog (Zeeman 1976) whose behaviour modes are attack and flight, although the control variables, fear and anger would seem difficult to observe independently.

It should be noted that the two control variables should be independent, in order to produce the cusp. If one variable simply represented costs of alternative A and the other costs of alternative B, that would amount to a single control variable, for instance $(A - B)$ or A/B , and one could simply state that x varies proportionately with $(A - B)$, or for complete bimodality that $x = 1$ when $A > B$, and $x = -1$ when $A < B$, yielding a step function.

An alternative way of defining the control variables is to define a *normal factor* (n on figure 3) and a *splitting factor* (s on figure 3) which in this case could be regarded as equivalent to the vector sums $(q - p)$ and $(q + p)$ respectively. In general the normal factor represents the forces tending to push the state variable to one extreme or the other, while the *splitting factor* determines whether the change is smooth, on the single sheeted part of the surface (low s) or discontinuous, on the double sheeted part of the surface (high s).

A good example of a cusp surface representing bimodal behaviour is afforded by Euler buckling (Zeeman 1976), and here the force acting transverse to the beam is the normal factor, tending to

make it go either up or down, while the compression on the beam is the splitting factor, enhancing the stability of equilibrium whichever position it is in and resisting change.

We can take a different and informative pair of control variables governing a bimodal choice to illustrate normal and splitting factors rather than conflicting factors. We take x once again as the state variable representing a two-way strategy mix, for instance the number of gas and electrical appliances in use in the house of the individual. Change to a new value of x is induced by change in the normal factor n which measures the economic benefits arising from the innovation. The splitting factor s represents the installation costs of the innovation, or the lack of capital for investment.

The argument follows much as before. If s is low, change in n will be accompanied smoothly by change in x , and the householder will shift from electricity to gas and back with no difficulty or delay. But if s is high, representing the high cost for changing the central heating (for example) from gas to electric, or vice versa, then we shall be over the two-sheeted part of the cusp (see Renfrew 1978). As the running costs of electric installation become progressively lower, there will be a sudden discontinuous switch as the individual previously using gas predominantly switches to electricity for all but a very few appliances.

The cusp (figure 3) shows a number of very interesting properties. That of *discontinuity* has already been discussed. Hysteresis (figure 5) is the phenomenon whereby the 'jump' does not take place for the same value of n , the normal factor, when the jump is from the upper to the lower sheet of the surface as when it is in the other direction. Following the Perfect Delay Rule (see Amson 1975, 182; Isnard and Zeeman 1976, 49), the system acts so as *locally* to optimise utility: so long as there is a maximum value for x , the state will not jump discontinuously. In other words, although there is another, more distant value of x with a maximum which would actually give a greater value of ϕ , the utility function, no jump will occur until the fold in the sheet is reached, and the total maximum for the present value of x entirely disappears. An alternative behaviour is described as Maxwell's Convention (Amson, *ibid.*; Isnard and Zeeman, *ibid.*), whereby the value of x moves at once to give the global maximum (the dotted line in figure 5) rather than optimising locally until the local maximum disappears. Other features have been well described by Zeeman (1976) and by Amson (1975, fig. 8).

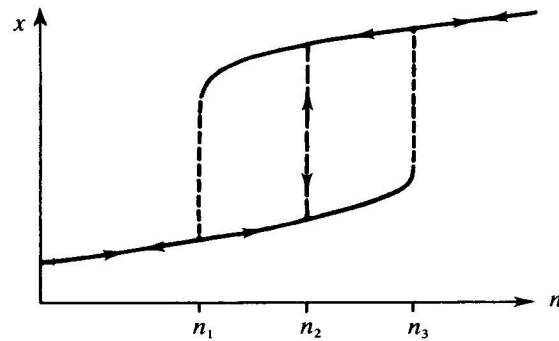


Figure 5. The hysteresis effect. Change in the behaviour variable x against that of the normal factor n (for a constant positive value of the splitting factor s). The perfect delay rule gives a hysteresis effect, while Maxwell's convention (represented by the broken line in the centre) does not.

Bimodal Choice and Group Behaviour

So far we have considered the form which individual choice follows, in the face of two options. It does indeed appear to be the case that individuals generally show hysteresis in the exercise of bimodal choice, sticking to what is known rather than following Maxwell's Convention and moving at once to a new global optimum. One reason, of course, is that the individual can rarely be sure that he has an unobstructed and undistorted view of the competing maxima. In practice embarking on a new course of action (with discontinuous change in x , the state variable) may well entail taking risks. I would argue that the notion of 'risk' is equivalent, in the framework of the present model, to 'imperfect view of competing utility maxima'. The longer terminology may not at first sight seem an improvement – but it is because it brings the entire discussion into relation with the unifying concept, the cusp.

A second reason is simply the difficulty of teaching an old dog new tricks. We have spoken of installation costs as a splitting factor, delaying innovation and inducing discontinuous behaviour. There is a direct psychological analogy here, where the time and mental energy involved in mastering a new process is the direct analogue of installation cost. It is easier not to change 'unless you have to' – i.e. to follow a strategy of local optimisation.

It is my contention that most innovations involve, at least initially, little more than new ways of doing old things. Even metallurgy

seems to have started in most areas with techniques suited to the working of stone (drilling and cold-working before annealing), and with products (thick copper axes resembling stone ones, and small decorative objects) simply replacing existing goods. The first products of any new material are generally skeumorphs. It follows then that the initial decisions involved in the acceptance of invention were whether or not to carry out a particular process or produce a particular object, in the new way (with new material) or in the old way. All innovation therefore involves bimodal choice in the first instance. Only later do technical developments amplify the existing divergence so that important innovations are seen as new products and new technologies in their own right. Seventy years ago an automobile was simply a 'horseless carriage', and a radio a 'wireless' telegraph; the 'talkie' emerged from the 'movie', and the moving picture from the kinetoscope.

An individual can change behaviour mode either as soon as the new maximum is globally more attractive than the old (Maxwell's Convention), or at the last moment, when the old maximum disappears (Perfect Delay Rule), or somewhere in between (Imperfect Delay). There is no way he can change earlier than the first of these, or later than the second, without failing to maximise utility even locally. The Imperfect Delay convention is discussed by Amson (1975, 197), and several procedures are in fact possible. One is that there should be a certain threshold value, and once the global maximum exceeds the local maximum by more than that value, change will occur even though the local maximum has not yet disappeared.

Now let us consider a group of individuals whose past history has been much the same, and who are stationed at approximately the same point on the cusp surface. For the moment let us imagine them all following the Perfect Delay Rule. First if they are perfectly homogeneous – identical to one another – and subject to the same influences, they will all change at once, even though the behaviour of one does not affect the behaviour of another. Secondly, if there is now some variation in their behaviour, some dispersal in their position, they will not come together to the edge of the fold of the cusp, and they will not all change together. The overall effect will then be a fuzzing out of the discontinuity. But in the third case, if their behaviour is now in some way coupled, with positive feedback between individuals, so that the action of one affects that of others, they may well move together although not themselves perfectly homogeneous.

Discontinuity and Long-Term Change

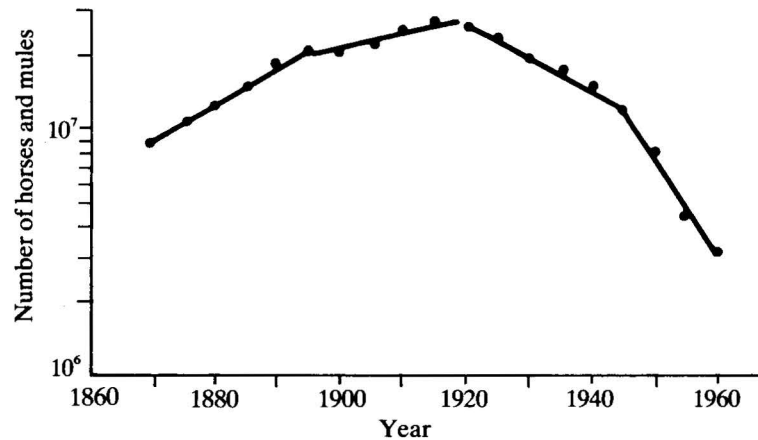


Figure 6. Gradual change in traction strategy (number of horses and mules on a logarithmic scale) after Hamblin, Jacobsen and Miller (1973). The curve is essentially a smooth one, artificially and needlessly represented in straight-line segments.

Homogeneous Individuals (No Coupling)

An elegant example of such behaviour is afforded by the growth and decline of the railroad system in the United States. It is one of the clearest examples I have found of catastrophic behaviour involving bimodal choice.

In their discussion of exponential growth and exponential decline, Hamblin, Jacobsen and Miller (1973, 92) graphically illustrate the rise and decline in the use of the horse in the US (figure 6). Although they link the points with straight line segments, supposedly describing 'exponential epochs' (since the vertical scale is logarithmic), a smooth curve might do as well. We may imagine this as described by the equation $N = e^{kt}$, where k is also a variable and changes smoothly and gradually from a positive value in 1870 to zero around 1920 to a subsequent negative value.

If we fit this onto our picture of the cusp catastrophe, with the exponent k as the state variable (represented by x in figure 3) then some normal factor is declining while the splitting factor is constant at zero. In effect we are saying that the exponent k is a function of t : $k = f(t)$. Indeed approximately $k = l - m \cdot t$ (where l and m are constants not varying with time). So far this has nothing to do with the bimodal behaviour of the cusp catastrophe: it can adequately be shown on the single-sheeted surface of the cusp.

The Anatomy of Innovation

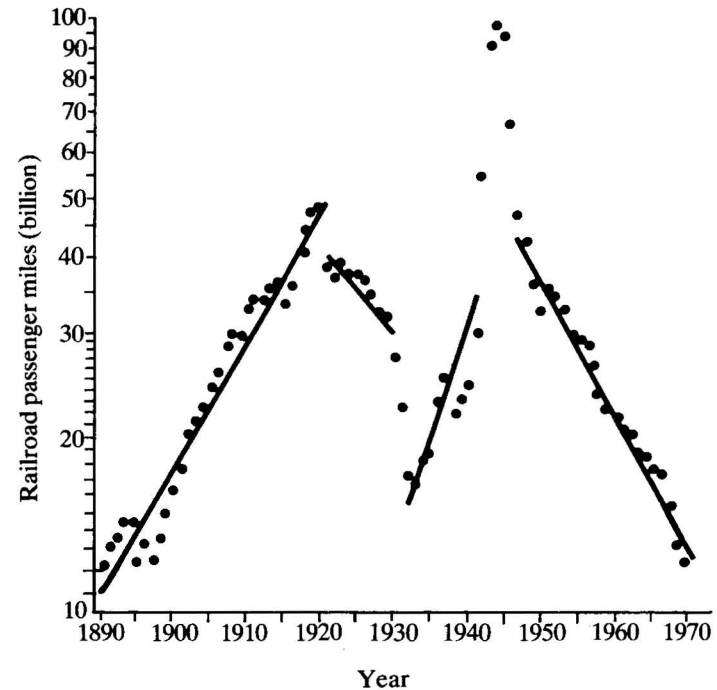


Figure 7. Catastrophic jumps in travel strategy, resulting from bimodal choice between road and rail travel under varying conditions of cost and convenience (railroad miles on a logarithmic scale) after Hamblin, Jacobsen and Miller (1973). Sudden changes of slope represent sudden changes of the exponent.

But turning now to figure 7 we see very different behaviour. Hamblin *et al.* (1973, 94) have drawn straight lines as shown, indicating that the exponent k , which is equivalent to the slope of the lines, when the ordinate is plotted on a logarithmic scale, is switching from a high positive value (1890–1920) to a high negative value (1920–1932), to a high positive again (1932–1942) further increasing from 1942 to 1947, with a catastrophic switch to a strong negative value from 1947 to the present.

Very reasonably they relate these changes to competition between rail and car:

The 1917–1925 period also ushered in a new epoch in automobile use, an epoch in which the automobile became the dominant mode of transportation, apparently because of its

improved reliability and lasting qualities and its differential convenience and privacy. (*ibid.*, 95)

Reversal from 1930 to 1942, with increase in railroad miles 'probably occurred because many people simply could not afford the luxury of automobiles during the depression' (*loc. cit.*). That trend was accentuated during World War II, and reversed in 1947, with increase in automobile mileage and decrease in rail mileage.

We can now use the cusp with k as the state variable to give behaviour closely analogous to that described in the last section, and for precisely the same reasons, with the same control variables. Sudden switches in the value of k (the slope of the lines in figure 7) arise with gradual change in the control variables.

We are seeing here, therefore, the operation of bimodal choice *en masse*. The constraints are acting equally upon the great bulk of the population, and are reflected in mass decisions. Many further examples could no doubt be found of such behaviour, not only in the field of transportation choice, but in advertising and commercial marketing, where sudden switches in consumer shopping behaviour can follow relatively modest changes in pricing. It should be noted that in this example only the switch in 1942, soon after the American entry into the War, was subjected to direct external constraint through rationing of petrol and tyres.

We cannot document any prehistoric strategy switches as clearly as this example, but effects of this kind, of purely local origin, may well underlie the sudden expansion of metallurgy in several regions in the early bronze age, or indeed the emergence of farming over a fairly wide area of the Near East after more than a million years of hunting and gathering.

Inhomogeneous Individuals (No Coupling)

If we now consider a group of individuals who are not perfectly homogeneous, although they have been subject to the same constraints (the same control variables) each will have his own utility function close but not identical with that seen in figure 3. For a given point (p, q) in the control plane there will be a whole series of values of x , perhaps normally distributed with a mean value x on the surface seen in figure 3. We are dealing now with a congeries of surfaces lying close to that seen in figure 3, but not identical to it.

If each individual follows the Perfect Delay convention, he will stay on the upper surface of his particular sheet as the control variables alter in such a way as to bring him to the fold and he will then switch to the lower surface. The effect, I assert, which will be

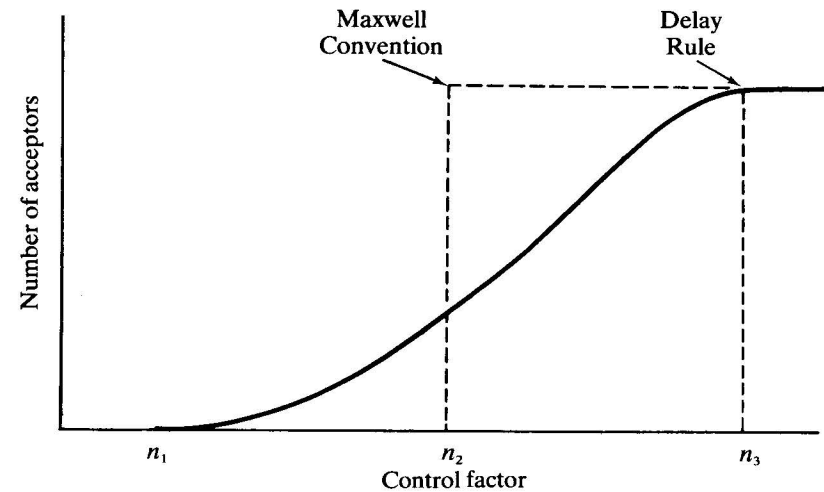


Figure 8. The number of acceptors plotted against change in the control factor (usually equivalent to change through time): (a) when individuals are homogeneous or if there is strong coupling between them – shown by broken lines for the alternative cases of Maxwell's convention and the perfect delay rule; (b) for inhomogeneous individuals with no coupling – shown by the bold line.

similar to the case where the entire population of points lies on the sheet seen in figure 3, but shows a slight fluctuation in the effective operation upon each of the control variables (p, q) , distributed normally about the mean individual with values (p', q') . The pattern of change will not then be the sudden switch (seen dotted in figure 8), which a single individual will still show, but a cumulative effect which corresponds closely with the 'adoption curves' shown in works dealing with innovation (Rogers 1962, 109).

If we show this instead, in figure 9, with rate of change on the ordinate, we shall simply obtain an approximation to the normal distribution. This can be arbitrarily divided into the 'innovators', 'early adopters', 'early majority', 'late majority', 'laggards' and 'non-adopters', of the conventional treatises on innovation (Rogers 1962, 110).

We can show, therefore, that such a pattern is predicted on the basis simply of a measure of inhomogeneity in the population. The effect would be accentuated if the inhomogeneity extended to the delay convention followed: in such cases those following Maxwell's

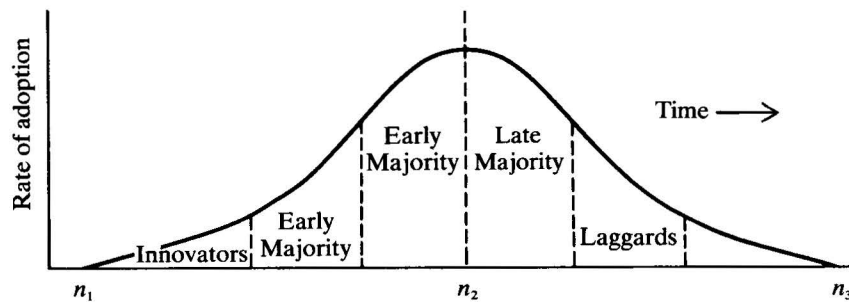


Figure 9. The 'diffusion' of innovation, after Rogers (1962). In reality nothing is gained by dividing this into 'innovators', 'early majority', etc. If the number of acceptors in figure 8 be designated N , the rate of adoption here is simply dN/dt , and no diffusion is implied.

Convention would be termed 'innovators', and those following the Perfect Delay Rule 'laggards'.

It should be noted however, that individuals do not make decisions about innovation in isolation, but are strongly affected by their neighbours and peers, so that there is a measure of coherence. The effect of this is to reduce the spread and promote bunching.

Coherence: Coupling Between Individuals

Human organisations – governments, business firms, societies – behave in certain ways like individuals. The analogy can be a misleading one, for instance when talking about the 'birth' and 'death' of a civilisation. Yet as concerns decision-making it has a certain usefulness. For it is (or used to be) the basis of good government, enshrined in the principle of cabinet responsibility, that the pros and cons of a decision are debated at length, after which a decision is taken, often by majority vote. The decision is then adopted as 'policy', and all party or government members, including those who opposed it warmly at a previous stage, are now expected to support it, or at least refrain from opposition. The observed behaviour is thus perfectly bimodal, and the fuzzing seen in the behaviour of inhomogeneous individuals without coupling does not appear. It is the feedbacks within the organisation which give it this property of coherence, of behaving 'as of one voice' like an individual, although in fact composed of many.

The behaviour of such organisations can be modelled by the cusp catastrophe, and indeed more complicated behaviour, including compromise, has been approached by Isnard and Zeeman (1976) in

terms of the butterfly catastrophe. The same state variable and control variables, seen in figure 3, are appropriate. The strategy mix implied by x will now represent company policy for the travel of its employees rather than the decision of a single individual.

When we are speaking of innovation, the argument is again much as set out above, with n , the normal factor, representing the advantages of change, for instance in reduced running cost, and s , the splitting factor, the capital (installation) costs of changeover. The behaviour of large organisations does indeed show discontinuity and hysteresis, as a consideration of the Concorde project will amply document. At the time of writing the British and French governments, following the Perfect Delay rule (local optimisation) must now be approaching the fold of the cusp, at which further production of the unfortunate aircraft will be stopped. Had they followed Maxwell's Convention (global optimisation), the project would have been scrapped as soon as it became clear that it was certain to lose money. In this case we are speaking of a decision or 'innovation' to stop producing a product, but the anatomy of the situation is no different when we are considering a positive decision to shift from hydrocarbon to atomic-powered electricity generation, or to computerise accounting.

Coherent group behaviour is modelled as effectively by the cusp catastrophe as is that of the individual. The analogy is important, yet it is equally important to see that it is simply an analogy. Group behaviour is *not* the same as individual behaviour: by using the same model to describe it we are focusing on those aspects which are analogous. Why the same model is appropriate to each is a more subtle and less obvious problem than it may at first seem.

An Example: the Inception of Metallurgy in Europe

The foregoing discussion offers us a framework within which to approach several specific and major innovations in the prehistoric period, of which the inception of metallurgy in south-east Europe will be considered here. This has traditionally been advanced as a classic case of the diffusion of culture (Childe 1957). The recognition of a number of relatively complex features including terracotta sculptures and copper objects at sites such as Vinča led to comparison with comparably complex traits at sites to the south and east such as Troy. The chronological arguments for believing the Balkan finds, such as those of Vinča, to be independent of the Aegean early bronze age have been advanced elsewhere (Renfrew 1969). So far the discussion has been polarised between those who

follow the view of Wertime (1964), quoted above on p.394, that metallurgy was invented only once, and those who argue for a number of independent metallurgical centres, in the Near East, in the Balkans, perhaps in Iberia (Renfrew 1967) and of course in south-east Asia as well as in the Americas. Instead, however, we can begin to discern that it was not the invention of the relevant techniques nor their dissemination in space, but their *adoption* in the cultural context in question which was the fundamental core of the innovation process. Moreover it becomes clear that a relatively sudden adoption is in some circumstances to be expected.

In several areas of the Old World an interesting pattern may be discerned, where many of the essential metallurgical processes were known at an early date, as occasional finds of copper objects testify. Yet the knowledge, the invention, was not widely exploited at this time: in the Balkans, in Anatolia, in Mesopotamia, in Iran and in Egypt we have, at different dates, the same pattern of an early knowledge of copper (the invention) and only later a rather sudden expansion of a metallurgical industry (the adoption).

In the Balkans, for instance, a copper awl 14.3 cm in length and made of native copper has been found at Balomir in Transylvania belonging to the end of the Körös culture and hence earlier than 5000 BC on a calibrated chronology (Bognár-Kutzián 1976). In the succeeding middle neolithic, around 5000 BC, copper beads are found in the Cernica cemetery of the Romanian Boian culture (Cantacuzino and Morintz 1963), and a whole series of beads, awls, chisels and fish-hooks are found in the late neolithic of the Balkans prior to 4500 BC. But it is only with the beginning of the so-called copper age of the Balkans that metal objects appear in great abundance, including solid hammer axes with their cast-in shaft holes (Charles 1967). One of these was found at Chotnitsa in Bulgaria in a Gumelnitsa culture context (Angelov 1958, 403). It is at this time that the Karbuna hoard with its 444 copper objects may be placed (Sergeev 1963). The find of small gold ornaments at Chotnitsa should also be noted. The relative chronology of these finds has been well discussed by Bognár-Kutzián (1972, 210–11).

If the early copper age sees a very marked development in the metallurgical industry, the late copper age, c. 4300 to 3700 BC sees its floruit. By this time copper mines were working at Rudna Glava in Yugoslavia (Jovanović 1971) and at Aibunar in Bulgaria. Several hoards of axe-hammers and chisels have been found at Pločnik in Yugoslavia. Most interesting of all are the cemeteries at Devnja, Goljamo Deltchevo and Varna in Bulgaria. At Varna a remarkable

series of gold objects has been found (cf. Gimbutas 1977), and the site rivals Troy in importance, the finds being perhaps a millennium earlier than the treasures of Troy II.

The point here, then, is that metallurgy began in a modest way at an early date in the Balkans, with the cold working and later the annealing of native copper. Only something like a millennium later was casting undertaken, but even the invention of casting did not at once produce a great upsurge in metallurgical production. That upsurge came, however, and it was associated both with new forms of useful tools (the hammer-axes and axe-adzes) and with a series of new prestige items. Significantly more than one metal was involved; for gold also played an important part.

All of this suggests that what needs analysing in each area is the *role* which metal objects played in the society before, during and after the phase of widespread adoption. If for heuristic purposes, the cusp catastrophe be selected as a model suitable for cases where bimodal choice may be involved, then the normal factor will be some measure of advantage in the use of metal objects. Advantage here implies efficiency in use, and takes into account both running cost and effectiveness. The splitting factor is some measure of the investment cost involved in producing or using metal objects. The value of the cusp configuration is that it shows us how gradual changes in these factors can bring about a sudden shift in the behaviour *x* of the individual or the group in relation to the use or non-use of metal objects.

It allows us to focus on adoption, the basic techniques (the invention) being available much earlier without promoting any large-scale industry. The widespread adoption of metalworking was relatively sudden when it came, although figure 9 reminds us that it was not instantaneous. Conceivably it may also have been reversible, like the changes in transport behaviour illustrated above (figure 7), although in most cases, as discussed below, the adoption of metallurgy ranks as an irreversible 'revolutionary' change. But in fact by 3000 BC metallurgy had become very much rarer or even extinct in the Balkans, and did not flourish again until the developed early Bronze Age of the area some centuries later.

No invocation of the cusp catastrophe will, in itself, result in a suitable analysis of the social and economic context of early metallurgy, but it does offer a framework. Clearly new inventions can have the effect of reducing the capital costs of existing production, so that in favourable cases newly invented techniques, whether originating inside or outside the society, can have instant effects: it

is not necessary to think of the cusp in such cases. But more significant will be changes in 'advantage' in using metal objects. And advantage includes not only practical ergonomic benefits, but any perceived advantage, however non-material. It is here that I feel the social aspects can most clearly be recognised.

In south-east Europe the marked increase in metal production in the copper age, which effectively determined the moment of adoption and hence of significant innovation, coincided with the widespread use of metal objects as remarkable prestige goods. During the neolithic there were already sporadic occurrences of metal objects as chisels, axes etc. – that is as tools of little importance, adding little to the range of functions already offered by stone tools. At the same time we see copper beads in cemeteries like Cernica, but in contexts where shell beads also occur (for instance *Spondylus* at Cernica). It is not until the next phase that gold pendants are seen (for instance at Chotnitsa) and the splendid gold plaques and other objects of the Varna cemetery. At this time also we begin to see the axe-hammers which are in fact the earliest metal shaft-hole tools, and are followed by the large axe-adzes of the later copper age.

I believe that we can begin to discern here the outlines of an explanation for the innovation. To argue it properly would require a careful analysis of the contexts of discovery of all metal objects of the period, and particularly of those contexts, such as burials, where the value or prestige-rating assigned to the objects might be estimated (cf. Shennan 1975). My prediction is that immediately prior to, and during, the real boom period – the catastrophe jump from little use to widespread use – metal objects will have taken on, for the first time, a high prestige value. This will then have shifted the value of the advantage factor sufficiently (and the shift need not have been a very large one) to bring the point on the equilibrium surface to the edge of the cusp.

In the Aegean, to take another region, an equivalent catastrophe jump is seen in the metal industry at the beginning of the Early Bronze 2 period around 2700 BC. Before that time copper was little used, although once again it is sporadically found in very much earlier contexts, (e.g. the copper axes from neolithic Knossos, from Sesklo and from Alepotrypa cave). Rather suddenly the metal industry changes altogether, and this change coincides with the first appearance of metal objects in cemeteries, with the first 'treasures' (at Troy, Poliochni etc.) and with the widespread use of the bronze dagger. The alloying of copper with tin to make bronze may have an important role here, but the boom appears to have happened before

alloying became universal, so that the use of tin (or arsenic) may be a consequence rather than a cause of the striking developments of the time. This important question of technological development and improvement is mentioned again in the next section.

The argument developed here, for the Balkans, in the sketchiest outline, may thus have a wider applicability. For, although the specific control factors would differ somewhat in each case, the basic cusp catastrophe configuration, relating to bimodal choice, could still hold. In each case, internal developments within the society would have shifted the 'advantage' factor, and these changes will have been as much social as technical. Similar arguments may apply for the north European bronze age. For there again, in some areas, copper objects had long been known, yet no major development occurred until the moment when metal adopted a striking prestige role, often in a 'Beaker Culture' context. Once again technical development (involving the use of bronze) followed rather than preceded the widespread acceptance of the metallurgy.

A similar sequence of events may have accompanied the widespread use of metallurgy in other parts of the world. Certainly the 'treasure horizon' in the Near East has often been the subject of comment: rich hoards and finds like those of Alaca Hüyük, Horoztepe, Troy, Poliochni and Ur seem surprising to us precisely because they come so early in the development of metallurgy in each area. The find of rich prestige goods at the beginning rather than the climax of the widespread use of metal in a region may be more than fortuitous.

The Great Revolutions

It was Childe who coined the terms 'Neolithic Revolution' and 'Urban Revolution', no doubt with the Industrial Revolution in mind, and in *The Bronze Age* (1930) gave reasons for thinking that in Europe, at least, the inception of metallurgy was of comparable significance. They were termed 'revolutions' because of the fundamental significance of the changes which took place, and Childe did not claim that they occurred rapidly in the original homeland. But he did stress in each case the rapidity of the adoption of the new process outside the area of its initial inception. In his view, the central organising process of European prehistory was diffusion, 'the irradiation of European barbarism by Oriental civilisation' (1958b, 70). The idea that metallurgy was diffused is still widely accepted, and the same view is widely argued for agriculture (e.g. Ammerman and Cavalli-Sforza 1973). Comparable arguments for

the spread of the urban way of life to the Indus and elsewhere were set out a few years ago by Wheeler (1961, 248), and are still accepted by many. I would argue that, in a very general sense, the same underlying pattern of change is operating in each case giving rise to analogies in the explanations offered. Of course each development in each area, to be properly understood, has to be studied in its own terms. But to the extent that the terms 'Metallurgical Revolution' or 'Neolithic Revolution' have any meaning, it may be valid to analyse their superficial similarity. Diffusion need no longer be the result of the analysis.

Barker (1977) has stated the position well:

The study of early metallurgy is developing along the same lines as the study of early agriculture; a choice between two apparently opposing models, diffusion or independent invention, is now seen to be a hopelessly inadequate theoretical basis for understanding when, how and why early agriculture developed in some parts of the world and not in others, and I would argue that the same is true for early metallurgy.

The key to an effective understanding in each case, I believe, is to separate the invention – the technical discovery, or the availability of a specific plant or animal species – from its widespread adoption. In the case of agriculture, sporadic finds of wheat or barley in upper palaeolithic deposits, not yet very well documented, suggest the possibility that these species may have been exploited in the Near East and perhaps in south-east Europe well before the neolithic period. The significant step would not then be the first use of the potential domesticate, but the shift to a pattern of exploitation with a very high dependence on a narrow range of species and a much less broad spectrum of exploitation in general.

If we now think of figure 3 as relating to the subsistence strategy of hunter-gatherers, with x representing the degree of dependence on cereal foods, then some insight can be obtained. The normal factor n represents the benefits of using cereals intensively, and s the capital costs of altering the subsistence strategy. In certain environmental or demographic circumstances a sudden shift of strategy must occur, and of course much thought has been devoted in recent years to analysing just what those circumstances might be.

There is one further ingredient in this case which makes the switch decisive, and in certain circumstances irreversible. It is that after some years or decades, if the new strategy is maintained, *new technological developments arise*. In the case of the cereal crops one of these is the formation of a tough rachis, which greatly increases

the effective crop yield. There are of course further changes associated with 'domestication' which make the domesticate a more satisfactory crop than its wild prototype. Much of the same is true of metallurgy: after a few decades of intensive exploitation (smelting and casting) of oxide ores in favourable areas, the use of alloys is likely to be discovered. Analogous consequences flow from the developments in craft specialisation associated with urban life. These are 'multiplier effect' phenomena (see pp.272–6 above) and in general they are not the product simply of human behaviour but also of the nature of the physical world. Only in such cases where a new strategy has unforeseen consequences of a technical nature do growth and change continue, in this way making the innovation irreversible. It is precisely to these cases that the term 'revolution' is sometimes applied.

[1978]

Bibliography

- Ammerman, A.J. and L.L. Cavalli-Sforza (1973) A population model for the diffusion of early farming in Europe, in C. Renfrew, ed., *The Explanation of Culture Change*, London, 343–58.
- Amson, J.C. (1975) Catastrophe theory: a contribution to the study of urban systems? *Environment and Planning B*, 2, 177–221.
- Angelov, N. (1958) Der Sidelungshügel bei Hotnica, in V. Beshevliev and V.L. Georgiev, eds., *Studia in Honorem Acad. D. Decev*, Sofia, Bulgaria Academia na Naukite, 389–403.
- Barker, G. (1977) Early metallurgy in the Troad: the European and Asian context. Unpublished paper presented at the conference on Troy, Sheffield, April, 1977.
- Bognár-Kutzián, I. (1972) *The Early Copper Age Tiszapolgar Culture in the Carpathian Basin*, Budapest, Akademiai Kiado.
- (1976) The origins of early copper processing in Europe, in J.V.S. Megaw, ed., *To Illustrate the Monuments: Essays in Honour of Stuart Piggott*, London, 70–6.
- Cantacuzino, G. and S. Morintz (1963) Die jungsteinzeitliche Funde in Cernica, *Dacia* 7, 27–89.
- Charles, J.A. (1967) Early arsenical bronzes – a metallurgical view, *American Journal of Archaeology*, 71, 21–6.
- Childe, V.G. (1930) *The Bronze Age*, Cambridge.
- (1956) *Piecing Together the Past*, London.
- (1957) *The Dawn of European Civilisation* (6th Edn.), London.
- (1958a) *The Prehistory of European Society*, Harmondsworth.
- (1958b) Retrospect, *Antiquity*, 32, 75–9.
- Clarke, D.L. (1968) *Analytical Archaeology*, London.
- Fellner, W. (1971) *The Economics of Technological Advance*, New York.
- Ferguson, J.A. (1976) Investment decisions and sudden changes in transport, *Surveyor*, 9th July, 1976, 10–11.